Article

# Student Risk Assessment for Identifying Needs and Evaluating Impacts

*Raymond E. Morley and James R. Veale*

Abstract: *The purpose of this article is to present a basis for evaluating the effectiveness of programs and services for at-risk children and youth. An instrument designed to identify most at risk to least at risk students is presented along with examples of how to assess and evaluate program effectiveness utilizing the instrument. The primary intent of applying the process described is to assist personnel to leverage resources for maximum benefit.*

The term "at risk" was originally defined in Iowa (Office of Educational Services for Children, Families and Communities, 1996) with the following results-oriented criteria: Children and youth (a) not meeting goals within ongoing education programs, (b) not completing high school, and (c) not becoming a productive worker upon leaving high school. Multiple criteria were identified in each of these three categories to assist in identification. A given student could be at risk by one or more of the three categories. The specific criteria used to identify students as at risk were drawn from a wide array of state and national information regarding factors that contribute to student failure and lack of success in school. Multiple criteria for identification are indicated and suggested for use in each of the categories. These criteria are still being used in Iowa schools to identify students who need additional assistance to succeed and to leverage resources to help students maximize success. These same criteria plus more are used in the enclosed risk assessment instrument intended to assist educators to identify at-risk children and youth, leverage resources, and assess the effectiveness of services provided. Multiple examples are provided to illustrate its utilization in the management and delivery of services and in assessing and evaluating their effectiveness.

## Student Risk Assessment Instrument

An instrument is presented in this paper for identifying students who are least at risk to those who are most at risk. This instrument was developed from team processing of program effectiveness by school and community-based support services personnel in the School-Based Youth Services Program in Iowa (Veale, Morley, & Erickson, 2002). In order to plan how to work together and make a difference for children and youth, team members needed to determine whether services were effective with the most at-risk children. Broad-based group data was not enough to demonstrate whether, in fact, the services were impacting those students most at risk. This previously hidden information was needed to develop the necessary knowledge to change services to help the most at-risk children and youth. The Student Risk Assessment Instrument moved the teams to more profound levels of knowledge for planning and leveraging resources.

The development and implementation of the instrument occurred from 1990 to 2000, a 10-year period of model program development between schools and multiple community-based support service agencies and organizations. Partial support for development came through the FINE (First in the Nation in Education) Foundation (Veale, 1995). The Student Risk Assessment Instrument serves as a tool to assist schools and school districts to determine the effectiveness of programs. Moreover, it allows observations of student performance on outcomes across risk levels, which can help with planning and modifying services, as well as resource management.

Thirty factors were identified by local community teams as significant reasons for students being at risk of not succeeding in school, dropping out of school, or not becoming a productive member of society. Seven factors were identified as "critical" for determining degree of risk, while the other 23 were considered important but "noncritical." A critical factor is one that may ***by itself*** force a student into a school failure, dropping out, or lack of productivity upon leaving school. The critical factors are (1) dropped out or expelled; (2) victim of physical, psychological, sexual abuse, rape or other violent crime; (3) pregnancy/teen parent; (4) homeless; (5) language/cultural barriers; (6) out-of-home placement; and (7) committed criminal acts. A noncritical factor is one which combined with other such factors (altogether, four or more) may force a student into school failure, dropping out, or lack of productivity. Noncritical factors include repeated school failure, no extracurricular activities, chronic health condition, gang member-

ship, and no identified career interests, *inter alia*. The Student Risk Assessment Instrument is presented in the Appendix.

The factors we came up with agreed closely with those established in the Phi Delta Kappa (PDK) "Study of Students At Risk" (Frymier, 1992a, 1992b). Although published a year before we developed our instrument, we were not aware of that study at the time. Since that study was based on data from more than 20,000 students, and all of the factors included in the resulting PDK template were associated with factors included in our instrument, we felt that this provided a degree of validity for the factors included in our instrument and their generalizability outside of Iowa.

Empirical data have provided further validation. For example, students classified as high risk were found to have higher dropout rates than those of medium or low risk. Since having previously dropped out of school is one of the factors contributing to risk, this result provides further evidence regarding the validity of risk assessment using this instrument. Reliability was assessed in a study where separate observers assessed the same students in a collaborative services program site in Iowa.

This instrument has been found to be useful in describing populations served, evaluating the impact of services in those populations, identifying student needs, establishing policy guidelines, and as a tool for leveraging resources for school improvement initiatives. The instrument has the following advantages:

- simple checklist format;
- three levels of risk assessment (low, medium, and high), allowing easy entry into the database and use in surveys via color-coding (e.g., for evaluating impact of services);
- validity based on comparisons with an instrument of established validity and empirical data;
- reliability based on indices of interobserver agreement and correlation;
- specifically targeted to students and families in collaborative services programs.

## Classification by Level of Student Risk

The classification by level of student risk is based on the number and types of factors identified for a student. A student is classified as having

- *low risk* if no factors were indicated;
- *medium risk* if one to three noncritical factors were indicated (no critical factors);
- *high risk* if (a) one or more of the critical factors were indicated or (b) four or more of the noncritical factors were indicated.

It is intended that staff members identify these risk factors for each student upon intake and update these assessments whenever risk increases significantly and new information becomes available on students. (If no information is available on a student, he or she is classified as having *unknown risk*. This may occur, for example, when a student has just entered the school or program.) The rationale for the above rule was (a) to provide greater weight to the critical factors, (b) incorporate a cumulative effect for the noncritical factors, and (c) insure practicality by keeping it simple to use.

For purposes of evaluating the impact of services, we suggest that *new information can increase—but should not decrease—the risk level of a student*. This does not mean that the student cannot overcome these risk factors. The only situation where a student's risk could decrease would be when the original assessment was in error. For example, suppose that a student's attendance for the year was incorrectly recorded as 85 days missed, whereas the actual number of days was 8.5 days—a transcription error involving a misplaced decimal point. This should not be confused with the situation where a student no longer indicates the risk factor(s), e.g., a student whose attendance had been very poor but who is now attending regularly. The risk factor (poor attendance) is still there; it is just not presently being manifested. In contrast, changing a student's risk classification from high to low (or medium) would reduce one's ability to demonstrate program impact using standardized measures or informal assessments. Since the focus of a demonstration is often those who are most at risk, there would be fewer records on which to make such an evaluation. In effect, this would be throwing away data.

We consider the level of student risk to be a background characteristic, not an outcome. As such, the assessment of student risk can yield a specification, restriction, or qualification of program effectiveness. Risk is not *itself* a measure of program effectiveness (outcome) in this system.

Professional judgment must be utilized and trusted in the application of this instrument. Local school personnel are given the flexibility to make the decisions on risk classifications of children based on available data outside the instrument itself. Information from multiple resources will be necessary in order to apply the classification of students most effectively. For example, information from human services personnel may be necessary to verify homelessness. [Note: A spreadsheet template is available to monitor and calculate the level of student risk, as well as summary statistics on the risk factors for the student population. This template may be obtained by request, free of charge, from the authors.]

## Assessment of Student Risk: What Do We Get From It?

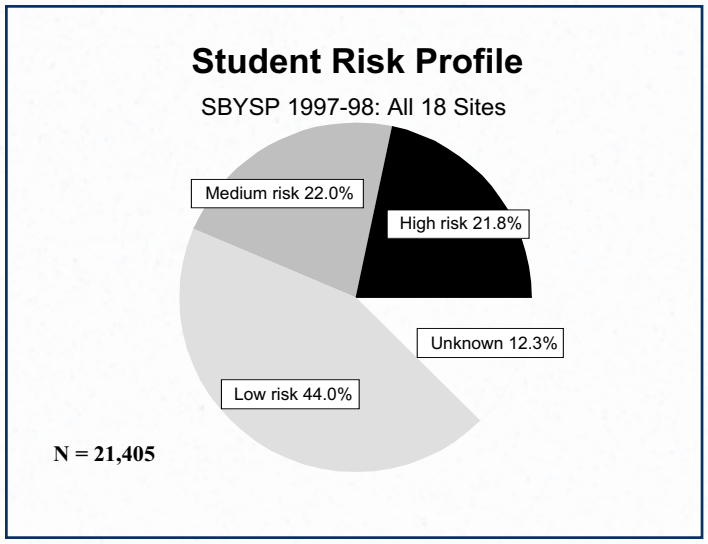
The assessment of student risk yields the following benefits for students, schools, and programs:

- *describing population served*—gives information on how many in a program are at high, medium, and low levels of risk;
- *identifying student needs*—provides a holistic, diagnostic picture of each individual student's needs (to personalize and fine-tune service delivery);
- *evaluating impacts of services*—determines the effectiveness of services for students at different levels of risk;
- *establishing policy guidelines*—determines the minimum number and type(s) of contacts for a student in a school year to increase the likelihood of positive outcomes (e.g., keeping the highest-risk students in school);
- *improving schools*—incorporates provisions for at-risk students, as identified by factors in the risk assessment instrument, into the comprehensive school improvement plan.

The first of these benefits provides an answer to the first part of the question that gave rise to the risk assessment instrument: "How do we know we are *serving* and *impacting* the most seriously at-risk students (in the community)?" We can determine the number of students participating in a program or initiative who are high, medium, and low risk. There may also be others in the community who are high risk and not participating in the support services program. If the instrument could be applied more generally to students, students not involved in services could also be assessed.

For example, a student's risk classification is included as a demographic variable ("risk factor") in the database *EASY/2EASY* used in the School-Based Youth Services Program (SBYSP) in Iowa (Veale, Morley, & Erickson, 2002). In the SBYSP in 1997-98, based on a total of 21,405 K-12 students served, we found that 21.8% were high risk, 22.0% medium risk, and 44.0% were low risk (12.3% were of unknown risk). This may be presented in a pie chart, as in Figure 1. In this example, slightly more than one student in five is high risk, and about half of those of known risk are either high or medium risk. Since the SBYSP is open to all students, these figures for high and medium risk may be considered fairly high. These figures will vary over program sites and over time.

**Figure 1. Student Risk Profile for the SBYSP in 1997-98**




The second benefit of student risk assessment is that it provides a holistic, diagnostic picture of the student's needs. This can be used in customizing or personalizing services and fine-tuning delivery of services. The value of using the risk assessment instrument as a diagnostic tool to drive service delivery is demonstrated by the following set of circumstances (Veale, Morley, & Erickson, 2002):

> A student is frequently absent, citing health problems as the reason. He is sent to the school nurse, and she learns that he has had frequent colds and other respiratory infections. The health symptoms are treated, but he continues to have health problems that cause him to be absent. The nurse becomes more concerned because she suspects that there may be other factors contributing to the student's health issues. She notes that the student does not appear to have warm clothing or a heavy winter coat. The nurse sets up a visit for the student with the school-based case manager who completes a more thorough assessment of needs with the student. As the case manager is assessing the various risk factors, he or she learns that the student is homeless and that he and his family are often forced to sleep in their car. Both parents dropped out of school before graduating and work part-time at minimum wage. They have no benefits such as insurance, sick leave, or vacation time. They cannot leave work to take their child to a doctor or clinic where they may have to wait several hours to be served.

The risk associated with being homeless is far greater than that of poor attendance and/or "colds" and alerts the case manager that a different type and intensity of services will be required.

Student risk assessment provides an opportunity to look for plausible relationships among many different variables and to gauge the type and level of intervention that may be necessary. Investigating many different factors also makes it more likely that the cause of the barriers to success can be discovered and addressed rather than focusing on an array of symptoms. In this case a cold would be a symptom of the student's more serious issue of homelessness. [1]

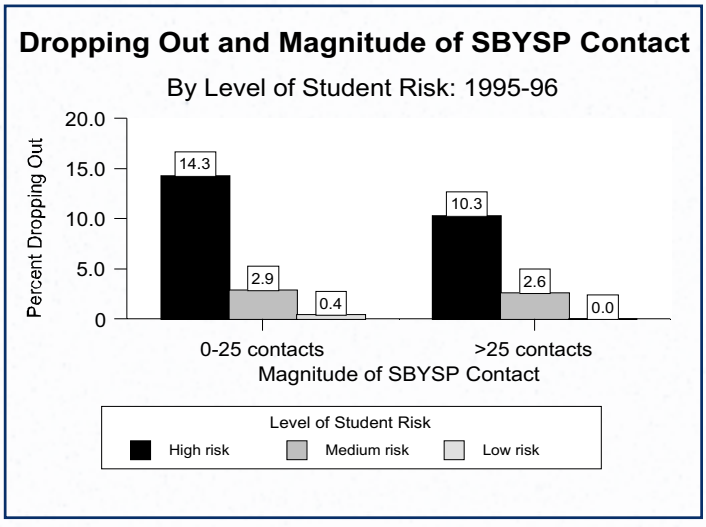
The third benefit—determining program impact for students at varying levels of risk—addresses the second aspect of the question that led to the development of the risk assessment tool. In outcomes evaluation, it is of interest to determine the degree to which performance on some outcome, for example absenteeism, is different for students at different risk levels. Such differences point to the importance of considering the social or cultural conditions (contexts) on which the impact of the initiative may be contingent (Pawson & Tilley, 1997). For example, if absenteeism is significantly reduced among the high-risk male student participants, this indicates that the initiative is contributing to improved attendance for male students most at risk. This result can lead one to question why the program isn't also successful with high-risk female students. Reflection and dialogue can result in changes in program focus or implementation that may yield significant improvement in attendance for all high-risk students.

Longitudinal analysis can add an important dimension to an evaluation. In the Caring Connection, a school-based collaborative services program in Marshalltown, Iowa, outcome data are added each year to the previous year's database. The premise here is that it is unrealistic to expect students to turn their academic lives around in one year. Multiyear data provide the opportunity to assess progress on outcomes over longer time intervals. For example, improvement in attendance is defined as missing no more than 10 days in the current school year after missing more than 10 days in the previous year. This definition may be applied to succeeding years to assess improvement over a longer time interval. In the Marshalltown program among high-risk students missing more than 10 school days in 1997-98, 17.9% improved in the following year and 26.3% (a 47% increase) improved in the third year—over their attendance in the first year. Among students missing more than 10 days in 1997-

98 who were medium risk, 23.3% improved in the following year and 37.7% (a 62% increase) improved in the third year, while among those who were low risk, 40% improved in the following year and 53.8% (a 35% increase) improved in the third year. This shows longer-term improvement among *all* risk categories, with somewhat greater percentage increases in improvements the third year (over those of the second year) among the medium-and high-risk students.[2] Moreover, among those of medium or high risk missing more than 10 days in 1997-98, the proportion improving their attendance from more than 10 days missed in 1998-99 to 10 days or fewer missed in 1999-2000 exceeded the proportion whose attendance worsened during this period ($P < .05$, McNemar test). This result implies that the longer-term improvement (over the three-year period 1997-98 to 1999-2000) was significantly greater than the short-term improvement (over the two-year period 1997-98 to 1998-99), for these higher risk students.

The fourth benefit is related to the third—establishing policy guidelines to increase the likelihood of success among students. For example, in the School-Based Youth Services Program in Iowa, it was found that high-risk students with more than 25 contacts with the program had significantly lower dropout rates than those with fewer contacts—10.3% compared with 14.3% ($P < .05$). This was not true for those at medium or low risk (see Figure 2). Thus, in terms of lowering dropout rates, the program appears to be impacting high-risk students more than those at lower-risk levels. Since high-risk students have the highest dropout rates, one of the program sites established a policy of encouraging high-risk participants to secure at least 25 service contacts with the program staff. Of course, the services must be appropriate to the specific needs of the student. The Student Risk Assessment Instrument provides the ability to fine-tune and personalize service delivery. Similar policy guidelines may be developed around other outcome areas or other types of programs.[3]

**Figure 2. Dropping Out and Magnitude of SBYSP Contact by Level of Student Risk**




All of the above discussions of benefits apply to school improvement initiatives. Local school districts are required under existing standards, largely driven by the No Child Left Behind legislation of 2001, to evaluate the effectiveness of existing programs and services for at-risk children and youth. The expectation is that this will occur at all levels of education (elementary through high school). Effectiveness of programs can be measured by identifying whether the most at-risk children and youth are improving and succeeding in school. It is important to consider the totality of risk factors and how these are distributed over the various groups mandated in the No Child Left Behind legislation to identify specific needs and achieve success (Foster, 2004).

Effectiveness can also be demonstrated longitudinally by a reduction in the percentage of children and youth who are at high risk (or an increase in the percentage who are at low risk). As stated earlier, this strategy has not been utilized in past research using the Student Risk Assessment Instrument, but the possibilities remain open for application in local school districts. In order to accomplish this type of measurement, some attempt would have to be made to reclassify students at given time periods such as (a) the grade levels for standardized testing, (b) September (the beginning of the school year) and May (the end of the school year), or (c) upon entry into school and upon exit or graduation.

Comprehensive school improvement plans identify evaluation strategies to assess student progress. Yearly progress reports are used to monitor the progress of students based on chosen procedures. The plans and progress reports can incorporate the above ideas to address at-risk children and youth including services provided and evaluation of effectiveness of those services. This data utilization would provide more precise assessment of progress with high-, medium-, and low-risk children and youth from a comprehensive point of view. Assessments could be conducted at the elementary, middle, and high school levels to evaluate effectiveness of services at each level, and resources could be leveraged accordingly. This system could also be applied at each grade level, if necessary, to identify program effectiveness and to leverage resources. In particular, federal and state funding sources identified in comprehensive school plans could be directed accordingly.

The case study on page 5 illustrates why we do *not* recommend erasing a risk factor, even though a student may no longer manifest the particular behavioral tendencies that define it. The fact that the student had those tendencies at one time means that he or she could return to them at some time in the future. We know, for example, that students who drop out are at increased risk of dropping out again. Moreover, although it may have receded, having the risk factor (e.g., poor attendance) at a previous time could make it more difficult to achieve outcomes during school or when the student gets out of school and into a work activity (e.g., showing up for work). The fact that, in some cases, students may overcome these risks and achieve success makes their story all the more impressive.

## Validity of Risk Assessment

The validity of an assessment is the quality of accurately assessing the desired construct, trait, or behavior. In this case, the construct is a student's risk—of dropping out of school, not success-

## Case Study

The following case study submitted by a local collaborative services program coordinator provides an example of how the Student Risk Assessment Instrument can help in organizing the various risk factors that are impacting the lives of students. It illustrates how the use of the tool is really a *process* that evolves as knowledge of the student's risk factors increases.

### *Example: Case Study Illustrating the Use of Risk Assessment to Diagnose Student Needs and Fine-Tune Service Delivery*

The risk assessment tool was initially used to determine if this particular student (we'll call him Bill) needed to be in a tutoring program. Four factors became apparent as we filled out the form:

- Experienced repeated school failure (Bill had failed several classes in the middle school);
- Poor attendance (his attendance had been sporadic for some time);
- No extracurricular school activities (he had not participated in any such activity);
- Economically disadvantaged (he was from a low-income family).

Four noncritical factors made him a high-risk student, and one for whom the tutoring program was appropriate. After being in the program for several weeks, it became apparent why he had been struggling in school. Bill opened up to me one day and told me about the physical abuse that he and his mother had been suffering at the hands of his father. Going back to the assessment tool helped us get a clearer picture of how at risk this young man was. We now had to add the following to his list of risk factors:

- Recent crisis or life transition (his father moved back into the home after having been gone for a couple of years);
- Extreme mobility (the family had moved several times to get away from the father);
- Victim of physical . . . abuse (the boy was a victim of physical child abuse by his father);
- Experienced mental health problems (we referred him to mental health counseling).

After the Department of Human Services became involved, things began to change. Some of the other risk factors faded as Dad moved away. However, new ones cropped up. Bill became the father in the family, taking care of a very mentally ill and depressed mother and two little sisters. We would have to add family dysfunction to the list of factors as he took on the parental role,[4] as well as substance abuse by a family member, as Mom was using (drugs). A new crisis appeared as Mom was placed into the Mental Health Institute. Sisters were removed and for a while Bill was basically homeless, with a neighbor taking care of him informally. This situation was eventually resolved.

As time has gone on, new factors have arisen. Bill has become sexually active; he has had relationship problems over a girl; and he committed a delinquent act (driving without a license). While some factors may be corrected or fade over time, their effects never seem to entirely disappear. For example, Dad may leave but the effects of the abuse continue to influence how Bill reacts to his environment. Attendance may no longer be a problem, but the effects of past poor attendance could influence his learning ability and future work attendance. Therefore, it is vital to never erase a risk factor but to look instead at their cumulative effect.

### *Strengths Indicated by Risk Factors Not Present*

Of particular interest are the factors that Bill has *not* experienced, which can be seen as strengths:

- He has stayed in school (no small accomplishment) and so has not become a dropout.
- His grades blossomed once he was no longer the caretaker of the family.
- He is healthy and does not appear to be using drugs or alcohol.
- He has personal goals and motivation to improve.
- He has not been involved with the juvenile court system (a delinquent act only gave him a ticket).
- He has the ability and desire to work.
- He has solid career plans.

With the support he now has, coupled with these strengths, we have a lot of hope for this young man.

fully completing a course of study, or not becoming a productive worker and citizen. Validity is often considered to be a characteristic of the instrument. Others consider validity to be a quality of the inferences or assessments based on a specific application of the instrument (McMillan, 2001). The latter is probably more accurate, but the language "validity of the instrument" is more common than that of "validity of the assessment." Moreover, validity is always a matter of degree. When quantifiable, this quality is often measured by indices or coefficients on a scale of zero to one (or zero to 100%).

1. *Content Validity: The Instrument Development Process*—Content validity refers to the extent to which the assessment items represent a larger domain of interest. Although theoretically quantifiable, this type of validity is usually in the form of a qualitative judgment. The process used to develop the instrument can contribute to this type of validity. In this case, the instrument was developed through a brainstorming process, with input from local program coordinators who were thought to be most knowledgeable about the types of problems students have in their families, school, or personal lives. A review process was used to further develop, fine-tune, and validate the instrument. These processes resulted in the factors identified in the risk assessment instrument. The emphasis was on the practical utility of the instrument—both in terms of the checklist format and the simple rule for classification. The authors and teams involved believe that this process resulted in a practical instrument that can be used to create a ***context*** within which to evaluate the effectiveness of local programs and services in reaching all children and youth, in particular the most at risk (Pawson & Tilley, 1997).

2. *Construct Validity: Agreement With Template Developed in Phi Delta Kappa Study*—Another approach to assessment validation is construct validity—how an assessment is related to an underlying construct, trait, or behavior, in this case, student risk. Often, construct validity is established by studying how an assessment is related to other assessments of the underlying trait. One such assessment is the "risk template" developed in a multiyear Phi Delta Kappa (PDK) study (Frymier, 1992b). A committee came up with 45 factors that previous research indicated contributed to putting children at risk. A protocol instrument was developed and experienced professionals in 276 schools in 85 communities collected data on more than 21,000 students in grades 4, 7, and 10 across the United States and Canada. Teachers or counselors who knew the students best and had immediate access to their records provided the information. These data were subjected to a variety of statistical and item analyses and the number of factors was eventually reduced to 24. These items were grouped into three categories: family, personal pain, and academic failure factors.

   There is considerable agreement between the 24-factor risk template developed in the PDK study and the 30-factor risk assessment instrument. For example, all seven of the critical factors in the risk assessment instrument are associated with those included in the PDK risk template. In some cases, there is a near perfect match (e.g., "pregnancy/teen parent" compares with "student involved in a pregnancy . . ."); in others, the critical factor in our instrument relates to factors in the PDK template (e.g., "homeless" relates to "mother or father . . . unemployed" and "student does not live with real mother and real father . . . "). Their classification criteria are also similar to ours—evidence of a single factor in the personal pain component (PDK) or critical factors (risk assessment instrument) was considered sufficient to assess the student to be seriously at risk. Evidence of two or more family factors and one or more academic factors was also considered sufficient to assess the student as seriously at risk using the PDK instrument. This criterion is comparable to that of four or more noncritical factors for identifying a student as high risk in the risk assessment instrument.

   A cross-correlation of factors indicates that all factors included in the PDK template are included or associated with those in the Student Risk Assessment Instrument, which includes other factors considered critical by Iowa educators. The Student Risk Assessment Instrument includes 12 factors not identified in the PDK final template, which bring it into close conformance with existing school standards. One may interpret this to mean that the factors included in our instrument are slightly more comprehensive, in order to align with existing standards for evaluation. The additional components relate to career development/education, which is identified as part of the education program of all students nationally (Secretary's Commission on Achieving Necessary Skills (SCANS), 1991, 1992). In addition, social factors were included to address the importance of human growth and development, also identified in current research as intrinsic to total student development (Adelman & Taylor, 2001; Goleman, 1995). Moreover, factors related to or leading to criminality were also included in the enclosed instrument (Catalano, 1999). The PDK factors included criminal acts, but not other factors leading to criminal acts such as gang membership and committing delinquent acts.

3. *Construct Validity: Correlations With GPA, Absenteeism, and Staying in School*—Another way to establish validity is by studying relationships between the assessments and other variables that are thought to be related (either positively or negatively) to the underlying construct, trait, or behavior. Three such variables are GPA, absenteeism, and (not) staying in school. Research indicates that at-risk students will tend to have lower GPAs, greater absenteeism, and reduced likelihood of staying in school (more likely to drop out). The first two are highly correlated with all other risk factors in the PDK study (Frymier, 1992a); the third includes being suspended or expelled from school, which is highly correlated with all other risk factors in the PDK study.

In Iowa's School-Based Youth Services Program in 1997-98, using the Student Risk Assessment Instrument and classification procedure presented earlier, the data on the three above-mentioned outcomes are presented in Table 1. Each relationship was in the

anticipated direction—decreasing GPA, increasing absenteeism, and increasing dropout rate for increased level of risk.[5] This provided additional evidence of the (construct) validity of the student risk assessment instrument.

## Interobserver Reliability: Agreement and Correlation Indices

The reliability of an assessment is the quality of consistently assessing the desired construct, trait, or behavior (student risk). Consistency can be defined internally or in terms of stability over time, forms, or observers. In this case, the most appropriate definition of reliability is consistency over observers. This is called interobserver (interrater, interscorer) reliability. Like validity, reliability is often measured by indices or coefficients on a scale of zero to one (zero to 100%).

In 2001, coordinators of the SUCCESS Program, a collaborative services school-based program in Des Moines, Iowa, agreed to participate in a study to assess the reliability of the assessments using the risk assessment instrument and classification procedure presented herein. The program case manager was asked to assess the risk levels of student participants in the program and, independently, have an individual from the school staff (counselor, teacher, etc.) assess the same students. Individuals who had knowledge of the students in question—their academic records, extracurricular involvement, and family situations—conducted the assessments.

Perhaps the simplest measure of reliability is the average proportion of matches, found by counting the number of factors on which the two observers agreed for each student, dividing by 30 (the total number of factors in the instrument), and averaging over the 108 students assessed. This yielded .835 or 83.5% matches on the factors indicated or not indicated. This may be broken down into separate proportions of matches for critical factors (.937 or 93.7%) and noncritical factors (.804 or 80.4%). These proportions may be the most appropriate measures of reliability for the diagnostic use of the instrument to customize and fine-tune service delivery.

The above results do not utilize the method of classifying students as high, medium, and low risk. The results incorporating this classification system are summarized in Table 2. Although this was not a random sample, the marginal totals are fairly typical of the risk distribution for this site. Note that these row and column totals reflect a higher level of risk than in the overall program for an earlier time period (cf. Figure 1).

The cells representing agreements between the case manager's assessment of the student's risk level and that of the school staff are shaded. The raw proportion of agreements is found by taking the total in these cells (81) and dividing by the total number of students assessed by both observers (108)—yielding 0.75 or 75%. This value indicated a fairly high level of interobserver agreement (McMillan, 2001).

Some of the 81 agreements could be due to chance. To correct for this, Cohen's kappa is sometimes used as an agreement index (Cohen, 1960). Expected values (based on the assumption of statistical independence between the two observers) were computed and subtracted from the numerator and denominator of the raw percent of agreements, yielding a kappa of 0.309. Although not large, this value is statistically significant (P = .0004).

The value of kappa is much smaller than the raw proportion of agreements. Given the marginal totals in Table 1, a high level of agreement between the two observers can be expected by chance alone. With the marginal totals given in this table, the ***maximum*** raw proportion of agreements is found by first pairing the marginal totals (in Table 2, (2, 7), (22, 17), and (84, 84)), taking the smaller of each pair (2, 17, and 84), summing (103), and dividing by the total sample size (108). This yields a maximum raw proportion of agreements of .954. Then correct for chance agreements as before, yielding a maximum kappa of .872. Another possible index is the ratio of kappa to its maximum value or "adjusted kappa"—0.309/0.872, or 0.354 (Traub, 1994). This doubly corrected agreement index has the advantage that it has a maximum value of one, which simplifies the interpretation.

Cohen's kappa counts only ***perfect*** agreements, that is, both observers assess the student at exactly the same level (low, medium, or high). This is a rather stringent criterion. For example, the 20 (= 9 + 11) who were assessed as medium by one rater and high by the other rater were counted as disagreements in computing kappa.

***Table 1***

**Outcomes GPA, absenteeism (more than 10 days missed), and dropout rate by level of risk among student sparticipating in the SBYSP in 1997-98.**

| Level of Risk | Outcome | | |
|---|---|---|---|
| | GPA | More than 10 Days Missed Per Year | Dropout Rate |
| Low | 2.67 (N = 3,794) | 27.0% (N = 8,428) | 0.4% (N = 5,156) |
| Medium | 2.23 (N = 2,403) | 38.4% (N = 3,644) | 2.8% (N = 1,934) |
| High | 1.89 (N = 2,155) | 52.9% (N = 3,061) | 13.2% (N = 2,374) |

*Table 2*

**Table of agreements between the case manager and school staff assessments of risk of students in the SUCCESS Program in 2001.**

| Level of Risk: Case Manager Assessment (#1) | Level of Risk: School Staff Assessment (#2) | | | Total |
|---|---|---|---|---|
| | Low | Medium | High | |
| Low | 1 | 1 | 0 | 2 |
| Medium | 4 | 7 | 11 | 22 |
| High | 2 | 9 | 73 | 84 |
| Total | 7 | 17 | 84 | 108 |



One might argue that some "partial credit" or weight should be given to such ratings. Weighted kappa using the "quadratic difference" weighting method, accomplishes this by assigning a weight of one to the diagonal cells (perfect agreement), to those that are just off the diagonal (near agreement: low on one, medium on the other or medium on one, high on the other), and zero to the two remaining cells (clear disagreement: low on one and high on the other) (Agresti, 1990). The weighted kappa is 0.451—a somewhat larger value reflecting the more liberal concept of agreement applied. It is also statistically significant (P = .0000). These indices of agreement utilizing the classification system may be the most appropriate for use of the instrument in evaluation. [6]

The various indices of interobserver reliability provide evidence of the consistency of assessments across different observers or raters using the risk assessment instrument. This is considered the most critical type of reliability for such assessments. [7]

Cohen's kappa counts only *perfect* agreements, that is, both observers assess the student at exactly the same level (low, medium, or high). This is a rather stringent criterion. For example, the 20 (= 9 + 11) who were assessed as medium by one rater and high by the other rater were counted as disagreements in computing kappa. One might argue that some "partial credit" or weight should be given to such ratings. Weighted kappa using the "quadratic difference" weighting method, accomplishes this by assigning a weight of one to the diagonal cells (perfect agreement), to those that are just off the diagonal (near agreement: low on one, medium on the other or medium on one, high on the other), and zero to the two remaining cells (clear disagreement: low on one and high on the other) (Agresti, 1990). The weighted kappa is 0.451—a somewhat larger value reflecting the more liberal concept of agreement applied. It is also statistically significant (P = .0000). These indices of agreement utilizing the classification system may be the most appropriate for use of the instrument in evaluation. [6]

The various indices of interobserver reliability provide evidence of the consistency of assessments across different observers or raters using the risk assessment instrument. This is considered the most critical type of reliability for such assessments. [7]

## Summary

The student risk assessment instrument presented in the Appendix has been found to be practical, valid, and reliable. It can help educators to (a) describe the risk levels in student populations, (b) diagnose student risk issues and fine-tune service delivery, (c) evaluate impacts of programs and services, (d) establish policy guidelines for programs and services, and (e) assist with school improvement and accountability initiatives. We offer this discussion not for the purpose of justifying a means to classify at-risk children and youth, but rather to support its use in managing and delivering services and in determining the effectiveness of such services. We recommend it to all who are concerned with assisting at-risk youth in their education and development.

## Endnotes

[1] "Homeless" is one of the demographic characteristics tracked in *EASY/2EASY*, a system for monitoring services and tracking student outcomes in collaborative services programs (see Veale, Morley, & Erickson, 2002). Homelessness is a factor that indicates high risk (in particular, high correlation with dropping out of school) in our instrument and is associated with at least two factors in the PDK template (Frymier, 1992a, 1992b).

[2] The numbers of students on which these percentages were based were as follows: high risk, 156 for 1998-99 and 133 for 1999-2000; medium risk, 86 for 1998-99 and 77 for 1999-2000; low risk, 15 for 1998-99 and 13 for 1999-2000. The slightly lower numbers for the 1999-2000 year reflects attrition due to dropouts, positive terminations (students successfully leaving the program), and/or missing data. Also, note the low numbers for the low risk students. This was due to the fact that we are focusing on those needing improvement based on attendance (missing more than 10 days in 1997-98), which is less likely for low risk students. Thus, the percentages for the low risk group are less precise than those for medium or high risk. Finally, the percentage increases in improvement were computed by dividing the percentage improvement for the third year by the percentage improvement for the second year, subtracting 1, and multiplying by 100.

[3] In 1997-98, data like those of Figure 2 were collected for the Caring Connection—the SBYSP site that established the aforementioned policy. The results were similar, with an even larger difference in dropout rates between the contact groups for high risk students in this site. To the extent that keeping students in school (their not dropping out) and improved attendance are related, this policy may have contributed to the positive result regarding long-term improvement in attendance among high- (and medium-) risk students in this program. (The Caring Connection was one of the four original SBYSP sites and was cited by researcher Joy Dryfoos as an outstanding "safe passage" program for youth (Dryfoos, 1998).)

[4] The family was dysfunctional before, but even more so now.

[5] The risk assessments using the Student Risk Assessment Instrument were made as part of the intake process (when the student entered the program) and, as more information was made available, adjusted (upward) as needed. The outcomes data cited in the table were collected at the end of the school year. Thus, the data in Table 1 may be considered evidence of ***predictive*** validity—the ability of the assessment to predict behavior or performance. However, since these outcomes have associated factors in the risk assessment instrument and some program sites may have reclassified students based on evidence of these outcomes (as well as other information) during the school year, there is probably some degree of functional dependence between level of risk and the outcomes cited.

[6] It may be argued that a risk factor ***score*** (equal to the number of risk factors indicated for the student) would have been a better indicator of the level of risk of the student. With this measure a correlation coefficient between the scores for the two observers would be an appropriate interobserver reliability index. In the reliability study, this correlation coefficient was found to be 0.601, which is statistically significant (P = .0000). Other possibilities include breaking this into a critical score (interobserver correlation of 0.740) and noncritical score (interobserver correlation of 0.587), as well as more sophisticated weighted scores (e.g., giving more weight to the critical factors). These were considered and rejected in favor of the simpler rule, which we felt had greater usability and practicality.

[7] For example, test-retest reliability is considered inappropriate here, since a student's level of risk can change over time. Inconsistent measures over time may occur due to actual changes in a student's risk profile—not measurement error.

## References


Adelman, H., & Taylor, L. (2001). ***Enhancing classroom approaches for addressing barriers to learning: Classroom-focused enabling***. Center for Mental Health Services, UCLA, Los Angeles, CA.

Agresti, A. (1990). ***Categorical data analysis***. New York: John Wiley & Sons.

Catalano, R. (1999). ***Positive youth development in the United States: Research findings on evaluations of positive youth development programs***. U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, National Institute for Child Health and Human Development.

Cohen, J. (1960). A coefficient of agreement for nominal scales. ***Education and Psychological Measurement, XX***(1), 37-46.

Dryfoos, J. (1998). ***Safe passage: Making it through adolescence in a risky society***. New York: Oxford University Press.

Foster, W. (2004). No Child Left Behind: Group at-risk composition and reading achievement. ***The Journal of At-Risk Issues, 10***(1), 1-6.

Frymier, J. (1992a). ***Growing up is risky business, and schools are not to blame***. Final report of Phi Delta Kappa study of students at risk (Volume I). Bloomington, IN: Phi Delta Kappa.

Frymier, J. (1992b). ***Assessing and predicting risk among students in school***. Final report of Phi Delta Kappa study of students at risk (Volume II). Bloomington, IN: Phi Delta Kappa.

Goleman, D. (1995). ***Emotional intelligence***: ***Why it can matter more than IQ***. New York: Bantam Books.

McMillan, J. (2001). ***Essential assessment concepts for teachers and administrators***. Thousand Oaks, CA: Corwin Press.

Office of Educational Services for Children, Families and Communities (1996). Guidelines for Serving At-Risk Students. Iowa Department of Education, Des Moines, IA.

Pawson, R., & Tilley, N. (1997). ***Realistic evaluation***. Thousand Oaks, CA: Sage.

Secretary's Commission on Achieving Necessary Skills (SCANS) (1991). ***What work requires of schools: A SCANS report for America 2000***. Washington, DC: U.S. Department of Labor.

Secretary's Commission on Achieving Necessary Skills (SCANS) (1992). ***Skills and tasks for jobs: A SCANS report for America 2000***. Washington, DC: U.S. Department of Labor.

Traub, R. (1994). ***Reliability for the social sciences: Theory and applications.*** Thousand Oaks, CA: Sage.

Veale, J., Morley, R., & Erickson, C. (2002). ***Practical evaluation for collaborative services: Goals, processes, tools, and reporting systems for school-based programs***. Thousand Oaks, CA: Corwin Press.

Veale, J. (1995). ***Developing research tools and evaluating the impact of multi-service school programs*** (Comprehensive narrative). Prepared for the FINE Education Research Foundation, with assistance from Morley, R., Erickson, E., & Graber, G., Des Moines, IA.

## Acknowledgments

The authors would like to thank Sharon Baughman, Cyndy Erickson, Donna Hempy, and Todd Redalen who assisted in the development of the instrument and contributed data or other materials included in this manuscript, and Margaret Jensen-Connet, Nancy Wells, and Des Moines school staff members who provided data on the reliability of the instrument presented here. In addition, the authors would like to thank Dr. Tom Deeter, David Winans, and Abby and Heather Morley for proofreading and clarification. Finally, the authors acknowledge the FINE Education Research Foundation for support of the research necessary for the development of this instrument and its application in determining the effectiveness of programs and services.

## Authors

**Raymond E. Morley**, Ed.D. is an adjunct faculty member at Drake University-Des Moines, Iowa. and is a consultant in the Iowa Department of Education responsible for dropout prevention programs, services for high school dropouts, at-risk programs, education of homeless children and youth, and school-based youth services programs. He has published over 50 manuscripts, including books, pamphlets, state guidelines and legislation, curriculum guides and journal articles. Ray.Morley@ed.state.ia.us;

**James R. Veale**, Ph.D. is a statistical/research consultant and educator in Des Moines, Iowa. He has authored/co-authored journal articles in statistics, applied mathematics, measurement, training and development, education, and medical research. His research interests include the costs of dropping out of school, evaluation of at-risk and prevention programs, and diagnostic educational measurement. j.veale@prodigy.net

## Appendix

### Student Risk Assessment Instrument

| CRITICAL FACTORS | CHECK (√) IF PRESENT |
|---|---|
| 1. Dropout or expelled | |
| 2. Victim of physical, psychological, sexual abuse, rape or other violent crime | |
| 3. Pregnancy/teen parent | |
| 4. Homeless (on the street, shelter, transitional housing, living with friends, or other temporary arrangement) | |
| 5. Language/cultural barriers | |
| 6. Out-of-home placement (foster care, detention, independent living, residential treatment, etc.) | |
| 7. Committed criminal acts | |
| NONCRITICAL FACTORS | |
| 1. Experienced repeated school failure (low achievement, low grades) | |
| 2. Poor attendance, repeated suspensions, repeated tardiness | |
| 3. Special education student or student with mental, learning, behavioral, or physical disabilities whose needs are not met through special education | |
| 4. No extracurricular school activities | |
| 5. Experienced mental health problems (including suicidal ideation or attempts or sudden personality changes) | |
| 6. Recent crisis (death, divorce, illness) or life transition | |
| 7. Social isolation/relationship problems/negative peer influence | |
| 8. Eating disorders | |
| 9. Chronic health condition | |
| 10. Sexually active | |
| 11. Substance abuse by self or family member | |

**Student Risk Assessment Instrument (Continued)**

| NONCRITICAL FACTORS (Cont'd.) | CHECK (√) IF PRESENT |
|---|---|
| 12. Economically disadvantaged | |
| 13. Lack of personal educational goals | |
| 14. Lack of motivation to improve | |
| 15. Family dysfunction/youth's needs are not being met by the family | |
| 16. Committed delinquent acts | |
| 17. Gang memberhsip | |
| 18. Extreme mobility (moving two or more times in one year) | |
| 19. Inability to keep employment/unacceptable work behavior | |
| 20. Lack of skills for competitive employment | |
| 21. No identified career interests | |
| 22. Lack of work ethic (not wanting to work) | |
| 23. No postsecondary work or training plan or goals | |